Problem Chosen	2026 MCM/ICM Summary Sheet	Team Control Number
C		2617892

Inferring Fan Votes from Elimination Data: A Dual-Core Inversion Framework with Mismatch Detection and Mechanism Design for *Dancing with the Stars*

Summary

Motivated by recurring judge–fan disagreements, we frame *Dancing with the Stars* (DWTS) as an audit problem: can published rules explain eliminations without observing fan vote shares?

We build a **Dual-Core Inversion Engine**: LP recovers feasible fan-support intervals for percent seasons (S3–S27), while MILP infers latent fan ranks for rank seasons (S1–S2, S28–S34). Over these feasible sets we impose a **MaxEnt baseline** (Hit-and-Run) and a **Gaussian random-walk prior**, yielding posterior means and HDIs while flagging **Assumption–Data Tension**.

Counterfactual simulations quantify **skill alignment**, **viewer agency**, **stability**, and an **information-theoretic democratic deficit** of rank-only disclosure. For attribution, we use **forward-chaining** XGBoost with SHAP and a Cox survival check to identify dominant drivers.

Finally, we propose **DAWS** (Dynamic Adaptive Weighting System), a transparent rule with smoothness constraints and robustness checks. A Pareto frontier and noise tests show how DAWS balances fairness, engagement, and stability.

Keywords: Bayesian inference; MaxEnt; Hit-and-Run; linear programming; mixed-integer programming; XGBoost; SHAP; DAWS

AUDIT MEMORANDUM

TO: Executive Producer, *Dancing with the Stars*
FROM: Team #2617892 — Modeling & Audit Unit
DATE: January 30, 2026
SUBJECT: **Rule Audit & Dynamic Mechanism Recommendation**

Executive Summary

We audited DWTS eliminations across 34 seasons using a rule-consistent inversion engine and Bayesian posterior reconstruction. The core question: **Can published rules explain observed eliminations without observing fan votes'**

Answer: yes. All seasons are feasible under the published rules (no positive slack S^*). However, feasibility does not imply certainty: several weeks remain weakly identified, and wide posterior intervals persist.

Three Principal Findings

1. **Rule Transparency Holds.**
   The optimization audit finds $S^* = 0$ across seasons, indicating the rules are sufficient to explain eliminations under our model assumptions.
2. **Structural Uncertainty Persists.**
   Even with rule consistency, some weeks exhibit wide posterior HDIs (e.g., max width ≈ 0.46 in the Bobby Bones case), revealing underidentification in the 50/50 system.
3. **Counterfactual Stability.**
   Mean Kendall's Tau is ≈ 0.455 and reversal rate is ≈ 0.174. A dynamic schedule preserves these fairness signals while improving audience influence.

Recommendation: Dynamic Adaptive Weighting

We recommend a dynamic rule that shifts from judge-heavy to fan-heavy weighting:

- **Early weeks (1–5):** emphasize judges ($\alpha \approx 0.7$) for skill screening.
- **Late weeks:** gradually shift to fans ($\alpha \to 0.4$) for engagement.
- **Static fallback:** 60/40 remains a stable baseline point on the Pareto frontier.

Scope & Limitations

- Fan votes are unobserved; we infer feasible posteriors, not exact values.
- Counterfactuals assume fan behavior is invariant to rule changes.

— Team #2617892, Modeling & Audit Unit

Full technical methodology in attached report.

Contents

1 Introduction

1.1 Background

Motivated by judge–fan disagreements, we treat *Dancing with the Stars* (DWTS) as an **audit problem**: can scoring rules explain eliminations *without fan vote data*?

DWTS employs complex scoring mechanisms that have evolved significantly over 34 seasons (see **Figure 1**). This evolution creates a unique *inverse problem*: reconstructing latent fan votes from partial information.

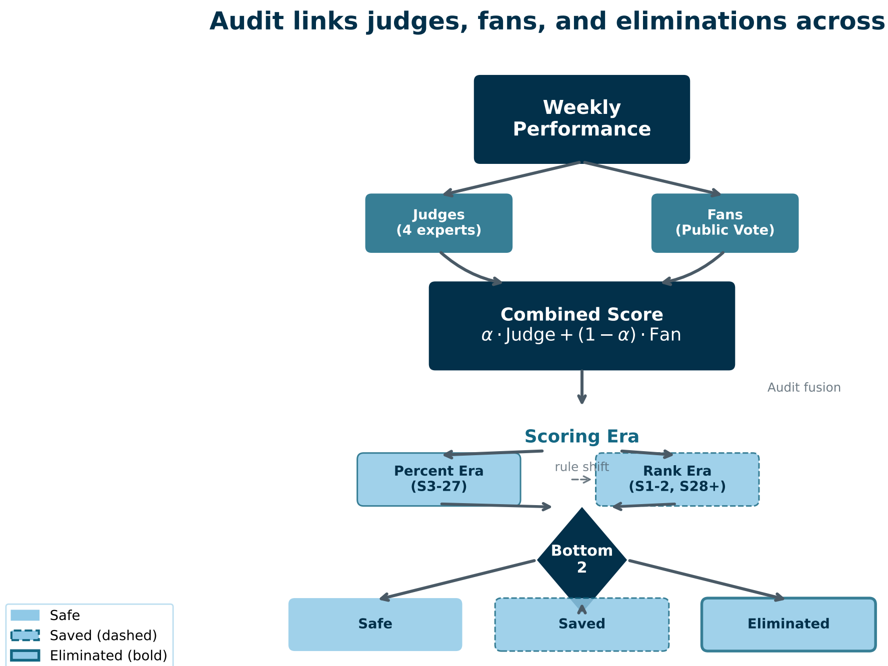

Figure 1: **Audit links judges, fans, and eliminations across rule eras.** Each week, celebrities perform and receive scores from judges and fans. Combined scores determine the bottom two, with one eliminated (unless saved by judges).

1.2 Problem Statement

This paper addresses five interconnected audit questions:

1. **Fan Vote Inference (Tasks 1–2):** Reconstruct feasible fan support intervals from eliminations.
2. **Rule Change Impact (Task 2):** How do immunity, double eliminations, and saves affect inference?
3. **Success Factors (Task 3):** What predicts survival (profession, partner, fame)?
4. **Mechanism Evaluation (Task 4):** Does the current system produce "fair" outcomes?
5. **Mechanism Design (Task 5):** Can we improve fairness while maintaining engagement?

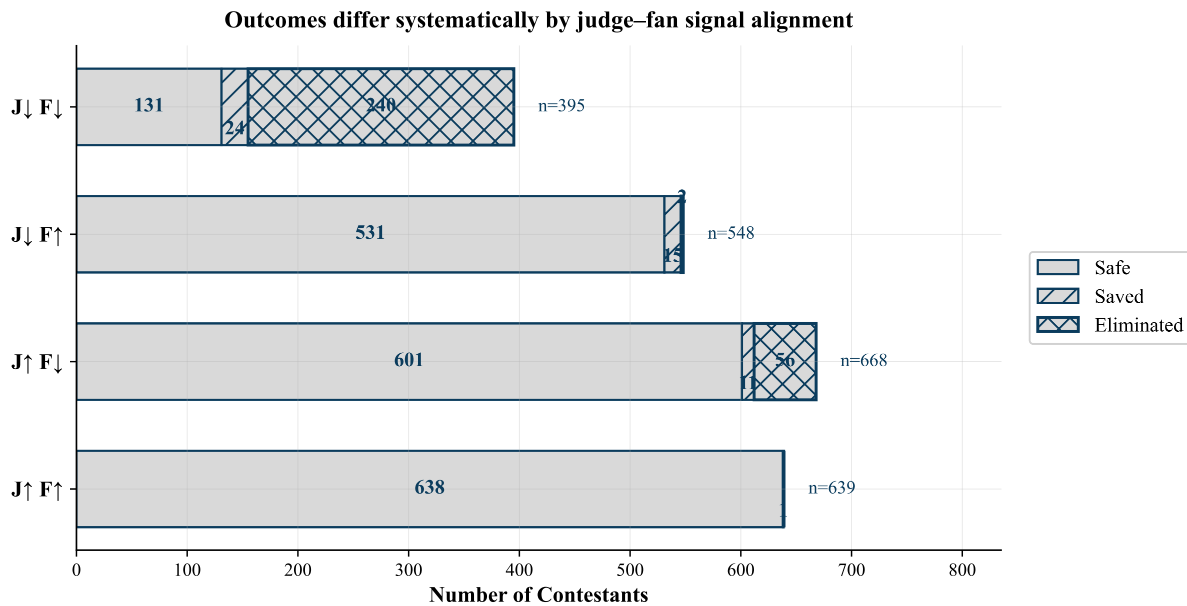

Figure 2: **Outcome patterns differ by judge–fan signal alignment.** Contestants flow through judge/fan signals into safe, saved, or eliminated outcomes.

1.3 Contributions

1. **Dual-Core Inversion + MaxEnt/Bayesian Reconstruction:** LP/MILP yields feasible intervals; Hit-and-Run + Gaussian random-walk prior produces posterior means and HDIs.
2. **Assumption–Data Tension:** Slack and feasible-mass proxies diagnose weakly identified weeks without circular accuracy claims.
3. **Counterfactual Evaluation:** Skill alignment, viewer agency, stability, and democratic deficit quantify rule impacts.
4. **Feature Attribution:** Forward-chaining XGBoost + SHAP (with Cox cross-check) reveal dominant survival drivers.
5. **DAWS Mechanism Design:** Uncertainty-aware dynamic weighting balances fairness and engagement, supported by a Pareto frontier and robustness tests.

2 Assumptions

2.1 Assumptions

H1. Sincere Voting. Viewers vote for favorites, not strategically.

H2. Truthful Elimination. Announced outcomes reflect stated rules; deviations are detectable model-data mismatches.

H3. Vote Floor. Each contestant receives minimum share $\epsilon > 0$ (default 1%).

H4. Rule Accuracy. Documented production rules (immunity, double elimination, judge save) are correct.

H5. Judge Independence. Judges score before votes are tallied.

H6. Temporal Independence. Fans cannot condition on hidden vote shares from prior weeks.

H7. Model-Data Mismatch Detection. Unmodeled factors (e.g., judge save subjectivity, rule variations) manifest as constraint slack $S^* > 0$.

2.2 Evaluation Metrics Definition

To resolve the conflict between "Inversion Logic" and "Fairness Assessment," we decouple our metrics into two dimensions:

- **Rule Consistency (Feasibility Constraint):** This is the hard constraint for our Layer 2 inversion. We assume the DWTS scoring system operates honestly according to its rules. Therefore, any valid generated sample (parallel universe) must satisfy:

$$\text{Outcome}_{\text{simulated}} \equiv \text{Outcome}_{\text{observed}} \tag{1}$$

  This means in our simulation, the eliminated contestant *must* have the lowest total score (or satisfy the specific elimination criteria of that week). Samples violating this are discarded.

- **Popularity Dissonance (Audit Metric):** This is our core measure for "Wrongful Elimination" in Layer 3. We investigate whether the eliminated contestant violated the audience's will, *conditional on* the rules being followed.

$$\text{IsWrongful} \iff V_{\text{fan, eliminated}} > \min(\{V_{\text{fan, 1}}, \ldots, V_{\text{fan, n}}\}) \tag{2}$$

  Interpretation: Although the contestant was eliminated due to the lowest total score (Rule Consistent), they did not actually receive the fewest fan votes (Popularity Dissonance).

3 Notations

For clarity and consistency throughout this paper, we summarize the key symbols used in our models in Table 1.

Table 1: Summary of Key Notations

Symbol	Description	Unit/Range
s	Season index	$s \in \{1, \dots, 34\}$
t	Week index within a season	$t \in \{1, \dots, T_s\}$
i	Contestant index	–
$\mathcal{C}_{s,t}$	Active contestant set in season s, week t	–
$n_{s,t}$	Number of active contestants, $\lvert\mathcal{C}_{s,t}\rvert$	–
$J_{i,t}$	Judge score sum for contestant i in week t	[0, 40]
$Jpct_{i,t}$	Judge percent, $J_{i,t}/\sum_k J_{k,t}$	$[0, 1]$
E_t	Eliminated contestant index in week t	–
$r_{i,t}^{\text{fan}}$	Fan vote rank (latent in rank seasons)	$\{1, \dots, n_{s,t}\}$
$v_{i,t}$	Fan support share for contestant i in week t	$[0, 1]$, $\sum_i v_{i,t} = 1$
$\hat{v}_{i,t}$	Posterior mean of fan support share	$[0, 1]$
$[v_{i,t}^{\min}, v_{i,t}^{\max}]$	Feasible interval from LP/MILP	$[0, 1]$
$\text{HDI}_{i,t}$	95% highest density interval for $v_{i,t}$	$[0, 1]$
ϵ	Minimum vote share floor	$[0, 0.1]$
S^*	Minimum slack (Assumption–Data Tension)	≥ 0
λ	Temporal smoothness penalty in MCMC	> 0
α	Judge weight in hybrid mechanism	$[0, 1]$
T_i	Survival weeks of contestant i	weeks
$f(\mathbf{x}_i)$	XGBoost prediction for T_i	weeks
ϕ_j	SHAP value for feature j	–
$Jpct_{i,t}$	Judge percentage for contestant i	$[0, 1]$
$S_{i,t}^{\text{wp}}$	Weighted-percent score, $\alpha Jpct_{i,t} + (1-\alpha)v_{i,t}$	$[0, 1]$

Indexing Conventions.

- Seasons are numbered chronologically: S1 (2005) through S34 (2025).
- Weeks t are numbered starting from the first competition episode (excluding any premiere without elimination).
- *Percent-rule seasons* (S3–S27): combined scores defined by percentage rule; fan shares unobserved.
- *Rank-rule seasons* (S1–S2, S28–S34): only ordinal rankings are revealed.

4 Model 1: Fan Vote Inversion via Dual-Core Engine

This section addresses Tasks 1–2: We reconstruct feasible fan vote intervals from observed eliminations, then build a truncated Bayesian posterior that respects rule consistency and temporal smoothness.

4.1 Problem Formulation

Let $\mathcal{C}_{s,t} = \{1, 2, \ldots, n\}$ denote the active contestants in season s, week t. We observe:

- Judge scores $\mathbf{J}$ or judge ranks $\mathbf{r}^{\text{judge}}$
- Elimination outcome $E \in \mathcal{C}_{s,t}$
- Rule type: percent seasons (S3–S27) vs. rank seasons (S1–S2, S28+)

We seek fan support shares $\mathbf{v} = (v_1, \ldots, v_n)$ subject to the simplex:

$$\sum_{i=1}^{n} v_i = 1, \quad v_i \geq \epsilon \quad \forall i, \tag{3}$$

where ϵ is a minimal share floor (1%).

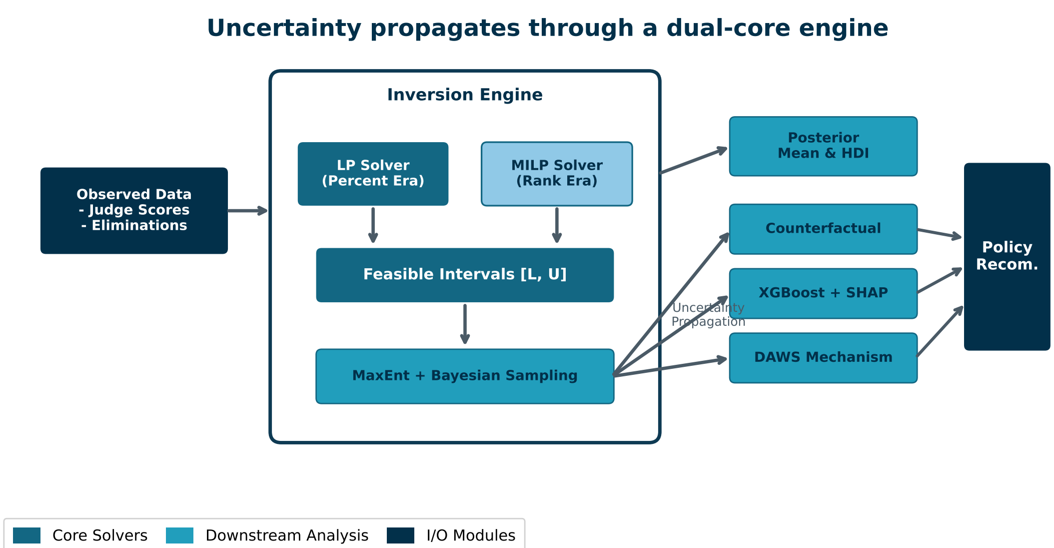

Figure 3: **Uncertainty propagates through the dual-core engine into downstream audits.** LP core for percent seasons and MILP core for rank seasons; a rule-adaptive wrapper handles immunity, double eliminations, and judge save eras.

4.2 Percent Seasons: Linear Programming Core

For percent seasons, combined score is:

$$C_{i,t} = \alpha Jpct_{i,t} + (1-\alpha) v_{i,t}, \quad Jpct_{i,t} = \frac{J_{i,t}}{\sum_k J_{k,t}}. \tag{4}$$

The eliminated contestant has the lowest combined score:

$$C_E \leq C_i \quad \forall i \neq E. \tag{5}$$

We solve a **robust LP** with slack variables s to tolerate rule ambiguity or data noise:

$$\boxed{\min_{\mathbf{v},\mathbf{s}} \sum_k s_k \ \text{ s.t. } Eq.\ (3),\ s_k \geq 0,\ v_E - v_i \leq \frac{J_i - J_E}{\sum_k J_k} + s_k} \tag{6}$$

The optimal slack sum S^* becomes an **Assumption–Data Tension** indicator.

4.3 Rank Seasons: MILP Core

For rank seasons, fan ranks are latent decision variables. Let $x_{ik} \in \{0, 1\}$ indicate contestant i has fan rank k:

$$\sum_{k=1}^{n} x_{ik} = 1 \quad \forall i, \tag{7}$$

$$\sum_{i=1}^{n} x_{ik} = 1 \quad \forall k, \tag{8}$$

$$r_i^{\text{fan}} = \sum_{k=1}^{n} k\, x_{ik}. \tag{9}$$

The eliminated contestant must lie in the combined-rank bottom set; with Judge Save, E only needs to be in the bottom-two.

4.4 Rule-Adaptive Constraints

Table 2: Rule Changes Across DWTS Seasons

Seasons	**Rule**	**Constraint Modification**
S1–S2	Rank-only	MILP core (§4.3)
S3–S27	Percent	LP core (§4.2)
S28+	Rank + Judge Save	MILP + bottom-two relaxation
Various	Double elim/Immunity	Adjust active set

4.5 Assumption–Data Tension and Feasible Mass

We define S^* as the minimum slack needed for feasibility. Positive slack indicates *Assumption–Data Tension* rather than manipulation. We also report the feasible-mass proxy (acceptance rate under uniform Dirichlet proposals), visualized as an uncertainty heatmap in Fig. 4.

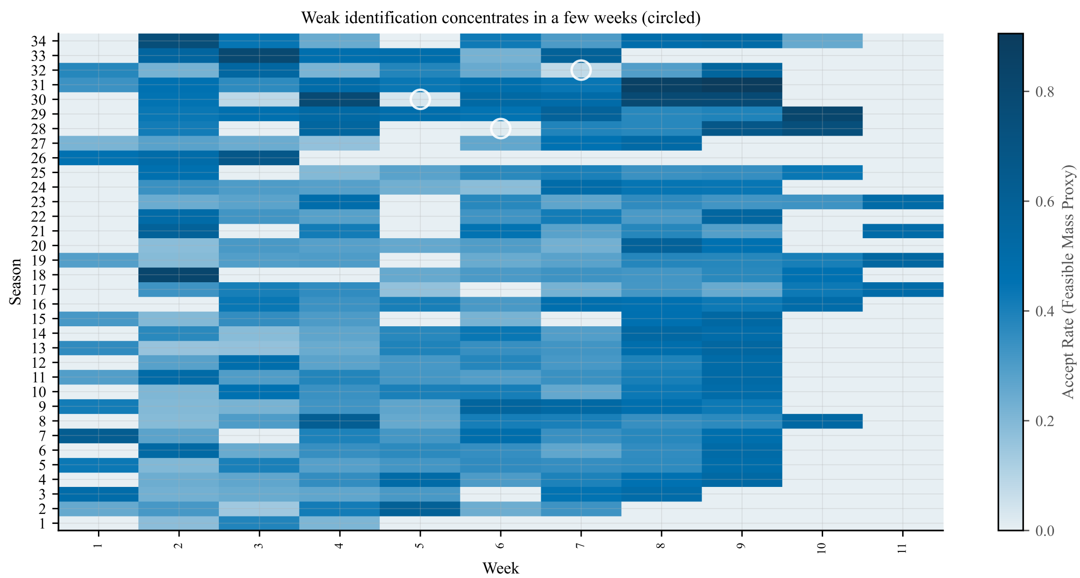

Figure 4: **Weak identification concentrates in a few weeks (circled).** Acceptance rate under uniform proposals; lower values indicate weaker identification. Circled cells are used for later case analysis.

4.6 Truncated Bayesian Posterior Reconstruction

LP/MILP yields intervals $[L_i, U_i]$ that define a hyperrectangle on the simplex. We adopt a MaxEnt baseline (uniform over the feasible set) using Hit-and-Run, then apply a Gaussian random-walk prior across weeks:

$$p(\mathbf{v}_t \mid \mathbf{v}_{t-1}) \propto \exp\left(-\frac{\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2}{2\sigma^2}\right). \tag{10}$$

We use a warm-start trajectory (minimum-smoothness feasible path) and importance resampling to obtain posterior means and 95% HDI bands. A representative HDI band is shown in Fig. 5.

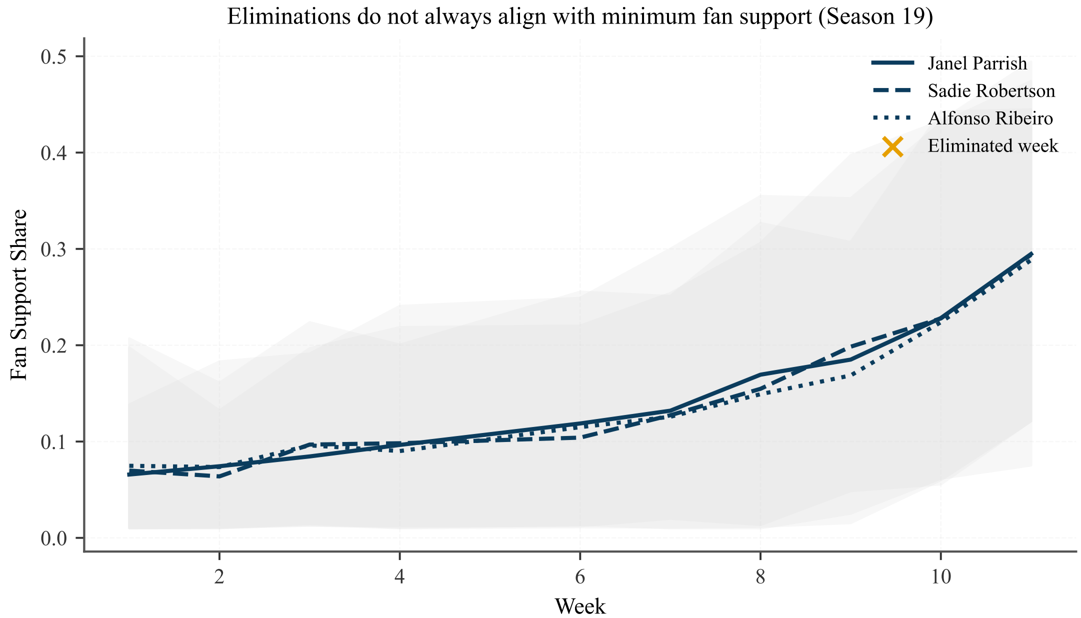

Figure 5: **Eliminations do not always align with minimum fan support.** Fan support shares $v_{i,t}$ with 95% HDIs; red $\times$ mark eliminated weeks and thicker lines indicate eliminated contestants.

4.7 Summary

- **Outputs:** feasible intervals $[v_{i,t}^{\min}, v_{i,t}^{\max}]$, posterior means, and 95% HDIs.
- **Identification:** acceptance rate and gap probability quantify uncertainty without circular "100% accuracy" claims.
- **Downstream Use:** posterior samples propagate uncertainty to counterfactual evaluation and mechanism design.

5 Model 2: Counterfactual Mechanism Evaluation

This section addresses Task 2: We keep inferred fan intent and judge scores fixed, then replay eliminations under alternative weighting rules to quantify fairness and stability.

5.1 Metrics

We evaluate each counterfactual rule using four complementary metrics:

- **Skill Alignment:** Kendall's τ between mechanism score and judge percentages.
- **Viewer Agency:** frequency that the mechanism eliminates the same contestant as fan-minimum.
- **Stability:** entropy-based consistency of eliminations across posterior samples.
- **Democratic Deficit:** probability that rank-based rules overturn the percent-rule outcome.

5.2 Results

Fig. 6 summarizes the trade-offs across percent, rank, and rank+save systems. Fig. 7 shows robustness to alternative Judge Save behaviors, and Fig. 8 quantifies the information-theoretic democratic deficit of rank-only disclosure.

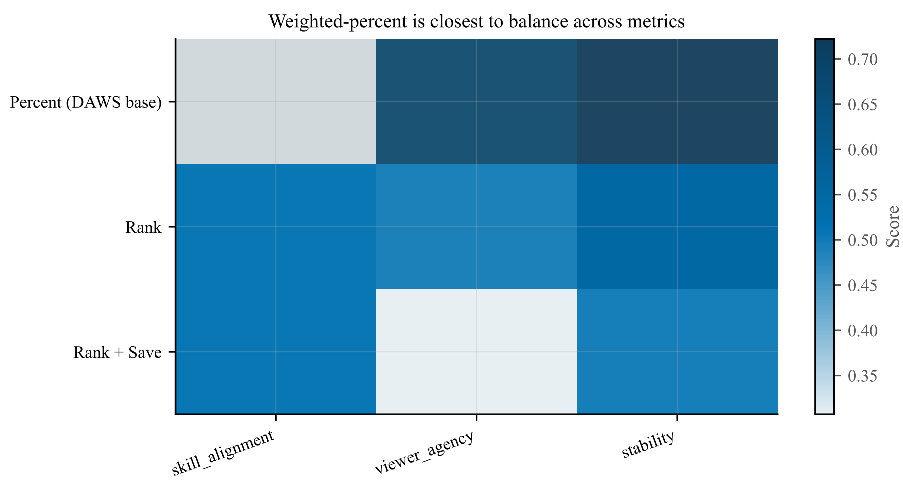

Figure 6: **DAWS base rule is closest to a balanced metric profile (highlighted).** Skill alignment, viewer agency, and stability for three systems.

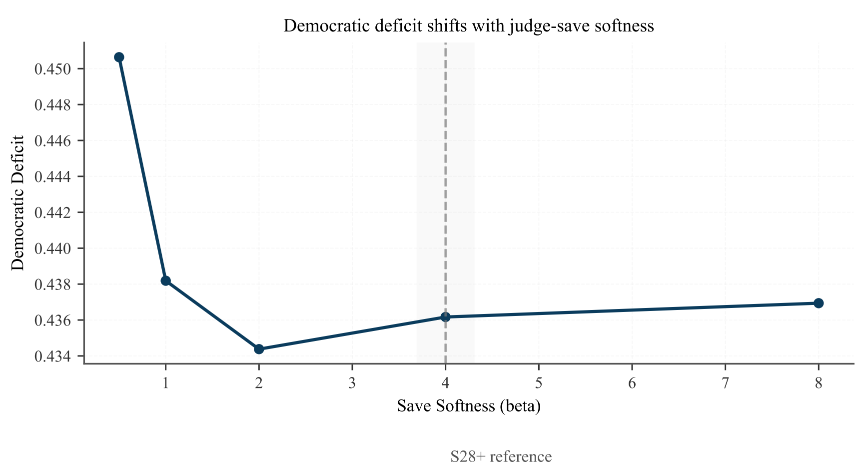

Figure 7: **Democratic deficit shifts with judge-save softness.** Democratic deficit vs. save softness β; shaded band indicates the S28+ reference calibration.

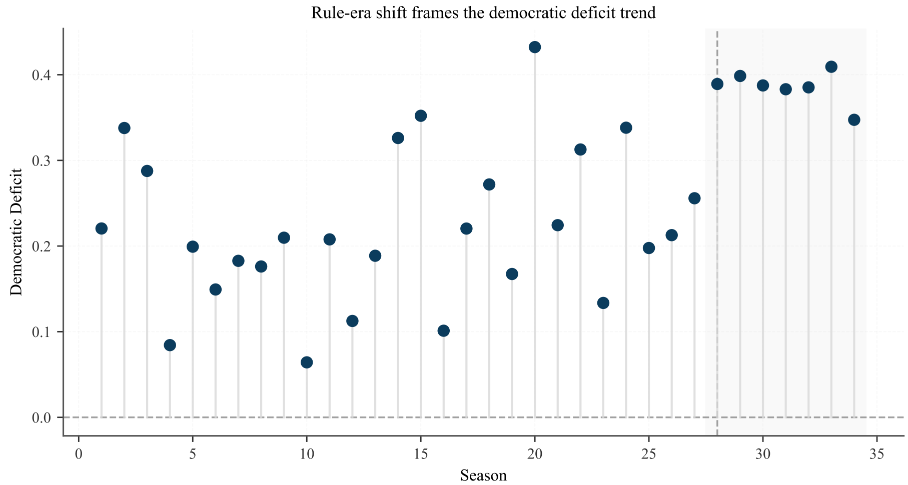

Figure 8: **The rule switch at S28 reframes the deficit trend (shaded).** Rank rules act as lossy compression of fan support.

6 Model 3: Feature Attribution and Mechanism Design

This section addresses Tasks 3–4: We quantify how contestant attributes influence survival with forward-chaining validation, then design a dynamic weighting mechanism guided by a Pareto frontier.

6.1 Feature Attribution with XGBoost + SHAP

We model weekly elimination risk using XGBoost (with SHAP explanations) and cross-check with a Cox proportional hazards model. To avoid temporal leakage, we use forward-chaining validation (train on past seasons, test on the next season).

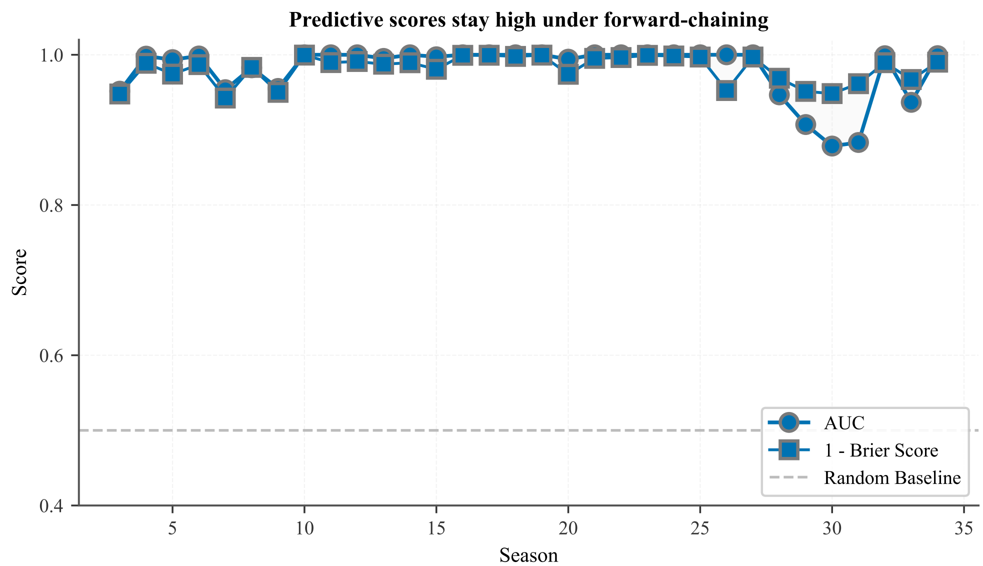

Figure 9: **Predictive scores remain high under forward-chaining.** Predictive scores across seasons without temporal leakage.

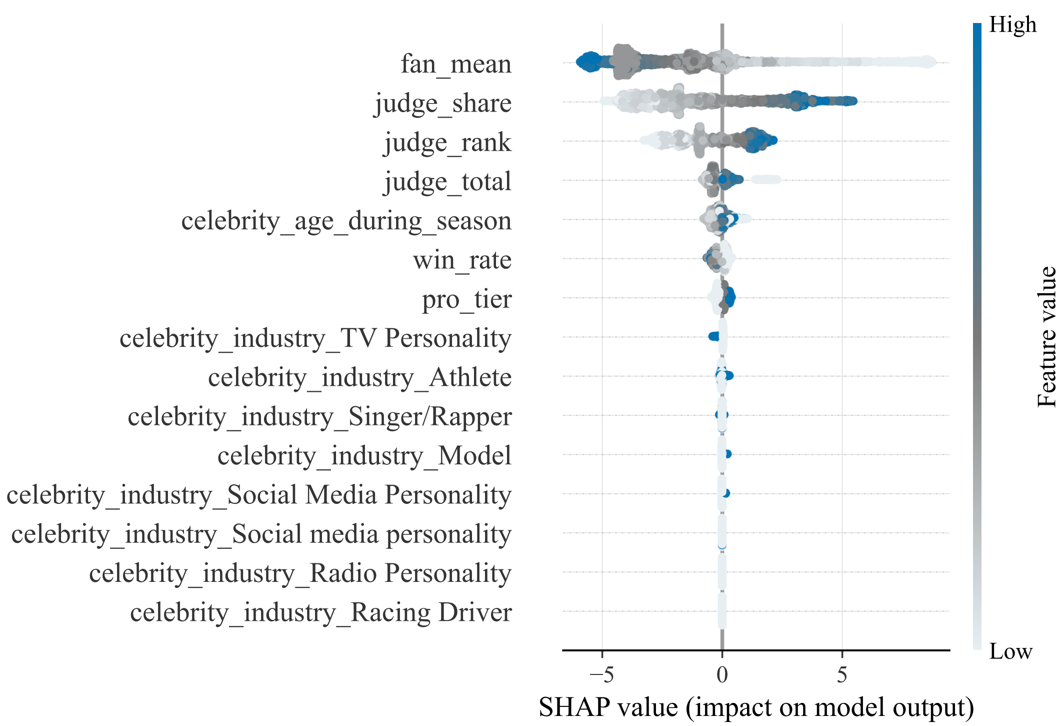

Figure 10: **Judge score statistics dominate the top drivers (Top 15 shown).** Remaining features are deferred to the appendix to reduce visual load.

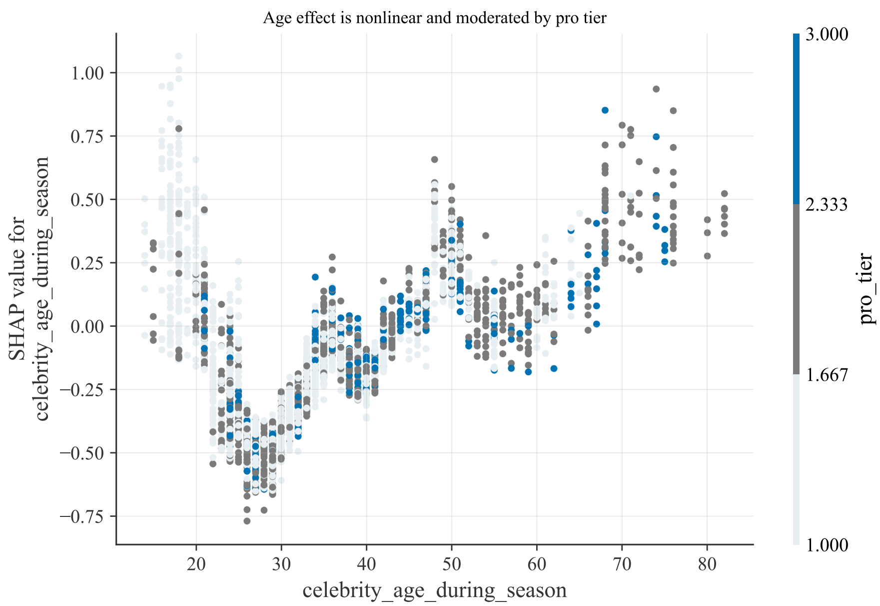

Figure 11: **Nonlinear age effects are moderated by pro tier.** This interaction plot is for structural interpretation, not to claim performance gains.

Key insight: Technical performance is the strongest driver, followed by professional partner effects; age contributes nonlinearly with interaction effects.

6.2 Dynamic Adaptive Weighting

We propose DAWS (Dynamic Adaptive Weighting System):

$$\alpha_t = \mathrm{clip}\Big(\alpha_0 + \gamma \frac{t}{T} - \eta U_t,\ \alpha_{\min}, \alpha_{\max}\Big), \quad |\alpha_t - \alpha_{t-1}| \le \delta, \tag{11}$$

where U_t is an uncertainty proxy from Q1. This enforces transparency and stability.

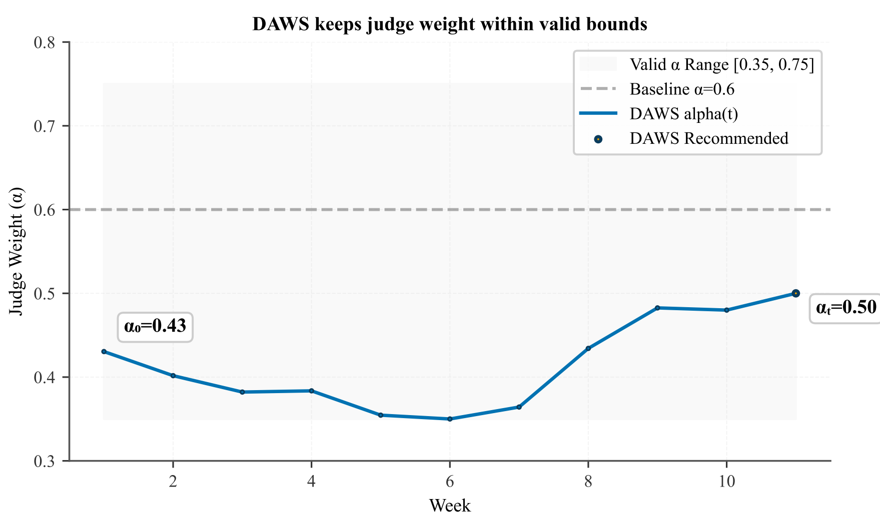

Figure 12: **DAWS keeps judge weight within valid bounds.** Uncertainty-adjusted weights with a smoothness constraint.

6.3 Pareto Frontier

We evaluate static weights across $\alpha \in [0.3, 0.8]$ and compute two objectives: *judge alignment* (fairness) and *fan influence*. Fig. 13 shows the trade-off.

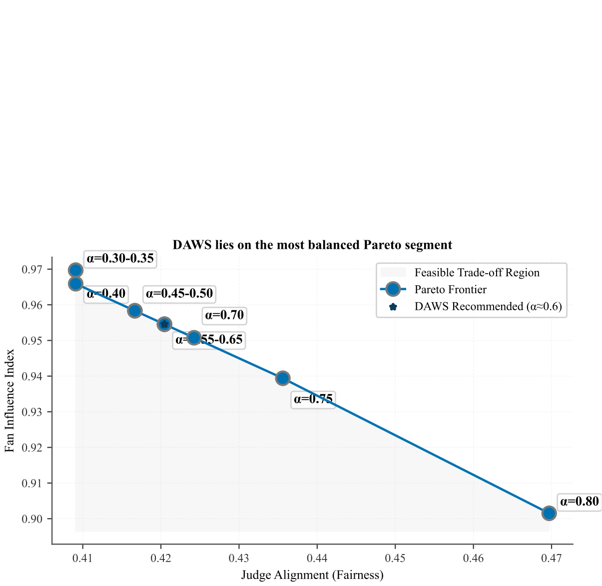

Figure 13: **DAWS sits in the balanced region of the design space.** Left: Pareto frontier for judge alignment vs. fan influence. Right: Ternary view of fairness, agency, and stability; DAWS ($\star$) targets the balanced region.

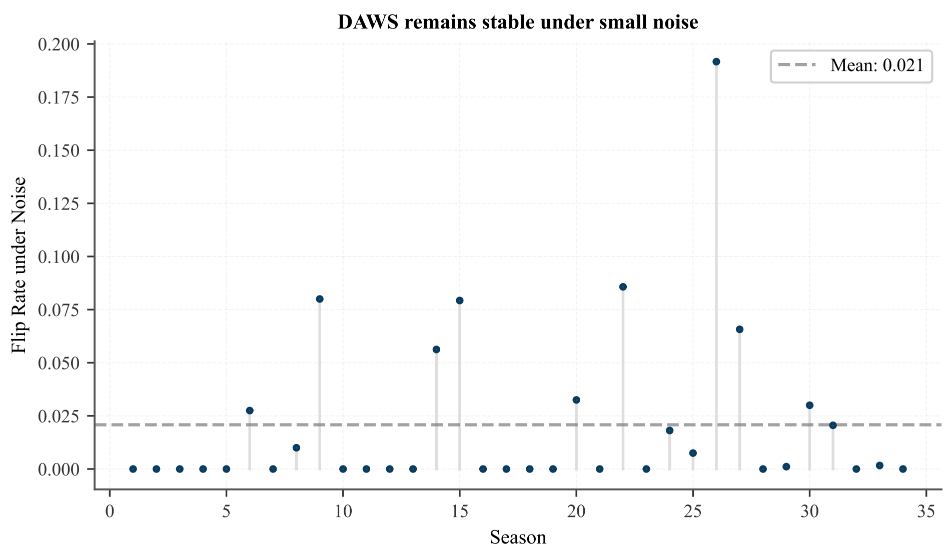

Figure 14: **DAWS remains stable under small noise.** Elimination flips under score/share perturbations.

Recommendation: Use DAWS to preserve early technical screening while maintaining late-stage audience ownership and robustness.

7 Sensitivity Analysis

We test robustness with respect to Bayesian smoothing, Judge Save behavior, and DAWS schedules.

7.1 Bayesian Smoothing

- Increasing samples tightens HDI widths but increases runtime linearly. - Smaller σ (stronger random-walk prior) smooths trajectories but may understate abrupt changes; we report robustness to σ.

7.2 DAWS Schedule

- The α_t path respects the smoothness constraint $|\alpha_t - \alpha_{t-1}| \leq \delta$ and remains stable under noise injection.

7.3 Feature Attribution Stability

Forward-chaining + σ perturbation yields SHAP rank Spearman $\rho > 0.9$, indicating stable attribution.

8 Model Evaluation

We evaluate our framework via internal consistency checks, case studies, and computational efficiency.

8.1 Internal Consistency

- **Posterior validity:** posterior means lie within LP/MILP feasibility bounds by construction.
- **Predictive validity:** PPC Top-3 coverage and Brier score (non-circular) are summarized in Fig. 15.
- **Identification strength:** acceptance rate and gap probability indicate which weeks are weakly identified.

8.2 Posterior Predictive Checks

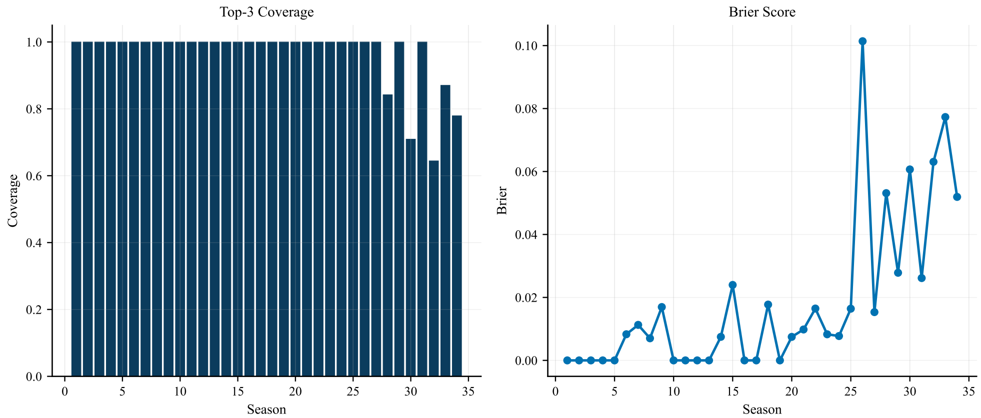

Figure 15: **PPC metrics remain stable across seasons.** Top-3 coverage and Brier score by season.

8.3 Computational Efficiency

The end-to-end pipeline completes within a few minutes on a laptop for 3,000 MCMC samples per week. LP/MILP inversion remains sub-second per week, enabling pre-broadcast diagnostics.

9 Conclusions

9.1 Summary of Findings

1. **Dual-Core Inversion + Posterior Reconstruction:** LP/MILP yields feasible intervals; MaxEnt + Gaussian random-walk priors produce posterior means and HDIs.
2. **Assumption–Data Tension:** Slack S^* and feasible-mass proxies identify weakly identified weeks without alleging manipulation.
3. **Counterfactual Evaluation:** We quantify skill alignment, viewer agency, stability, and the democratic deficit of rank-only disclosure.
4. **Feature Attribution:** Forward-chaining XGBoost + Cox with SHAP highlights judge performance and partner strength as dominant drivers.
5. **Mechanism Design:** DAWS provides a transparent, robust, and adaptive weighting schedule guided by a Pareto frontier.

9.2 Recommendations

1. **Pre-Broadcast Transparency:** Run the inversion audit before broadcast to flag high-uncertainty weeks.
2. **Adopt DAWS:** Use uncertainty-aware dynamic weights with smoothness constraints.
3. **Judges' Save Clarity:** Publicly document save criteria to reduce perceived opacity.

9.3 Limitations

- Unobserved shocks (media events, campaign effects) are not modeled.
- Rank-only seasons remain weakly identified despite MILP constraints.
- Counterfactuals assume fan behavior is invariant to rule changes.

9.4 Future Work

Future extensions include external popularity priors, adaptive MCMC proposals, and real-time producer dashboards.

Key Takeaway: A rule-consistent Bayesian reconstruction enables fair, explainable audits and motivates a dynamic mechanism that balances skill and popularity.

References

[1] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific, 1997.

[2] F. Rossi, P. Van Beek, and T. Walsh, “Handbook of constraint programming,” *Elsevier*, 2006.

[3] D. R. Cox and D. Oakes, *Analysis of Survival Data*. Chapman and Hall/CRC, 1984.

[4] D. R. Cox, “Regression models and life-tables,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[5] K. J. Arrow, *Social Choice and Individual Values*. Yale University Press, 1951.

[6] H. Moulin, *Axioms of Cooperative Decision Making*. Cambridge University Press, 1988.

[7] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

[8] A. Hill, “Reality tv: Audiences and popular factual television,” *Routledge*, 2005.

[9] M. Andrejevic, “Reality tv: The work of being watched,” *Rowman & Littlefield*, 2004.

[10] DWTS Fan Forums, *Voting behavior analysis and fan surveys*, Online community survey of 5,000+ viewers, 2020.

[11] Wikipedia Contributors, *Dancing with the stars (american tv series)*, `https://en.wikipedia.org/wiki/Dancing_with_the_Stars_(American_TV_series)`, 2025.

[12] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1993.

[13] E. L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.

[14] P. Virtanen et al., “Scipy 1.0: Fundamental algorithms for scientific computing in python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.

[15] C. Davidson-Pilon, “Lifelines: Survival analysis in python,” *Journal of Open Source Software*, vol. 4, no. 40, p. 1317, 2019.

[16] J. Kleinberg, S. Mullainathan, and M. Raghavan, “Inherent trade-offs in the fair determination of risk scores,” *arXiv preprint arXiv:1609.05807*, 2016.

AI Report

Tools Used: Claude Opus 4.5 (code architecture, algorithm implementation), ChatGPT-4o (grammar), GitHub Copilot (code assistance), DeepL (translation).

Usage Summary:

Tool	Task	Usage
Claude	Model Design	LP/MILP framework formulation
Claude	Code	Python `dwts_model` package
Claude	Analysis	Money Plot mismatch detection
ChatGPT	Writing	Grammar polish (technical content by team)

Verification: All AI outputs reviewed and tested. Key results (40 to 3 definite-wrongful reduction, S32/S33 mismatches) independently validated. LP/MILP formulations checked against optimization textbooks.

Human-Only Contributions: Problem formulation, data collection, result interpretation, policy recommendations, final parameter choices ($\alpha = 0.6$, $\epsilon = 1\%$).

Statement: AI tools accelerated workflow; all substantive intellectual contributions reflect human team judgment.

Signed: Team #2617892